THE FEASIBILITY AND VALIDITY OF PATIENT REPORTED OUTCOME MEASUREMENT INFORMATION SYSTEM (PROMIS) IN SYSTEMIC LUPUS ERYTHEMATOSUS

A Thesis

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Shanthini Kasturi

May 2016

ABSTRACT

Background: Accurate measurement of patient reported outcomes (PROs) is crucial to understanding how poor health impacts quality of life. PROs are particularly important in chronic diseases such as systemic lupus erythematosus (SLE), where disease manifestations are protean and objective measures may not capture patient centered domains. Patient Reported Outcome Measurement Information System (PROMIS) offers dynamic computerized adaptive tests (CATs) which have the potential to efficiently and accurately measure PROs that are relevant to SLE patients.

Objectives: The aims of this study were to: 1) assess the feasibility of administering PROMIS CATS to SLE outpatients; 2) assess the validity of PROMIS CATs by correlating them with legacy PRO measures; 3) correlate PROMIS CATs with standard measures of SLE disease activity and organ damage; 4) assess retest reliability of PROMIS CATs.

Methods: Adults meeting American College of Rheumatology SLE classification criteria were recruited from a SLE Center of Excellence. Subjects completed the Short Form-36 (SF-36), LupusQoL-US, and selected PROMIS CATs in domains of physical, mental and social health. SLE disease activity, flare, and damage were evaluated with the SELENA-SLEDAI and SLICC-ACR damage index. Subjects self-reported demographic information, relevant comorbid conditions, and their subjective experience completing the survey. PROMIS CATs were compared with disease activity, damage, and similar domains in legacy instruments using Spearman correlations (*r*). Retest

reliability was evaluated among subjects reporting stable SLE activity at two assessments one week apart using intraclass correlation coefficients (ICC).

Results: Of 238 outpatients approached, 204 (86%) completed at least one assessment, with 164 (80%) completing the assessment offsite. One hundred and sixty-two subjects (79%) completed a retest. There were no significant differences in demographic or clinical characteristics between those who completed the initial assessment and those who did not. Flaring patients completed the retest assessment less frequently ($p = 0.03$). Subjects found the questions relevant and validating. PROMIS CATs showed favorable performance characteristics and moderate to strong correlations with similar domains in both legacy instruments ($r = 0.49$ to 0.83, $p < 0.0001$). However, correlations between PROMIS CATs and the SELENA-SLEDAI and SLICC-ACR damage index were generally weak and statistically insignificant. PROMIS CAT retest ICCs were good to excellent, ranging from 0.72 to 0.88.

Conclusion: To our knowledge, these data are the first to show that PROMIS CATs can be successfully administered to a diverse cohort of SLE patients at the point of care or remotely, and are valid and reliable for many SLE relevant domains. Importantly, PROMIS scores did not correlate well with physician-derived measures. This disconnect between objective signs and symptoms and the subjective patient disease experience underscores the crucial need to integrate PROs into clinical care to ensure optimal disease management.

BIOGRAPHICAL SKETCH

Shanthini Kasturi, MD is a rheumatology fellow at Hospital for Special Surgery and New York Presbyterian Hospital-Weill Cornell Medicine and a candidate for the degree of Master of Science in Clinical Epidemiology and Health Service Research at Weill Cornell Graduate School of Medical Sciences. She completed her undergraduate studies at Harvard College in 2005, graduating magna cum laude with high honors in Social Studies and a certificate in health policy. She graduated with her medical degree from Harvard Medical School in 2010 and completed her internship and residency in internal medicine at Beth Israel Deaconess Medical Center in Boston, MA in 2013. Her research interests during fellowship have focused on the measurement of patient reported outcomes in systemic lupus erythematosus.

ACKNOWLEDGEMENTS

I would like express my deepest gratitude to my primary mentor Dr. Lisa Mandl for her enthusiasm and unwavering guidance and support throughout this project. I would like to thank my co-investigators Dr. Jessica Berman, Dr. Kyriakos Kirou, Dr. Alana Levine, and Dr. Lisa Sammaritano for critical support with patient recruitment, and Dr. Jayme Burket for assistance with statistical analyses. I would also like to thank Rima Abhyankar and Kelly McHugh for their diligent assistance with patient enrollment and the staff of the Hospital for Special Surgery Lupus and APS Center of Excellence for helping to facilitate patient recruitment. I am grateful to the faculty of the Masters Program, especially Dr. Carol Mancuso, and to my fellow students in the program for their comments and suggestions, which have helped shaped this project over the past two years.

Funding for this research was made possible by the Rheumatology Research Foundation Scientist Development Award.

TABLE OF CONTENTS

Biographical Sketch iii

Acknowledgements iv

Table of Contents v

List of Figures vi

List of Tables vii

Chapter 1: The Feasibility of Patient Reported Outcome Measurement Information System (PROMIS) in Systemic Lupus Erythematosus 1

- Introduction 2
- Methods 3
- Results 6
- Discussion 16
- References 21

Chapter 2: The Validity of Patient Reported Outcome Measurement Information System (PROMIS) in Systemic Lupus Erythematosus 24

- Introduction 25
- Methods 27
- Results 31
- Discussion 49
- References 55

LIST OF FIGURES

Figure 1.1	Study Flow Chart	7
Figure 1.2	Reminders Required for Initial Assessment	8
Figure 1.3	Reminders Required for Retest	14

LIST OF TABLES

Chapter 1

Table 1.1 Characteristics of Subjects Completing vs. Not Completing Initial and Retest Assessment 9

Table 1.2 Additional Characteristics of Participating Subjects 13

Chapter 2

Table 2.1 Characteristics of Participants 32

Table 2.2 PROMIS CAT and Legacy PRO Instrument Score Distributions 36

Table 2.3 Items and Time per Instrument 42

Table 2.4 Correlations Between PROMIS CATs and Legacy PRO Instruments 44

Table 2.5 Correlations Between PROMIS CATs and Physician Derived Measures 47

Table 2.6 Test-Retest Reliability of PROMIS CATs 48

CHAPTER ONE

The Feasibility of Patient Reported Outcome Measurement Information System (PROMIS) in Systemic Lupus Erythematosus

INTRODUCTION

The evaluation of patient reported outcomes (PROs) is a priority in systemic lupus erythematosus (SLE), a chronic autoimmune disease with involvement of multiple organ systems and significant impact on quality of life. PROs are recognized as an independent outcome measure in SLE, alongside disease activity and damage, and their assessment is required in clinical trials.[1–3] Historically, SLE PRO measures have been primarily designed for research purposes, but there is increasing interest in measuring PROs to enhance clinical care in SLE.[4,5] The routine capture of PROs can promote the provision of patient-centered care and performance improvement by facilitating patient-provider communication, informing treatment decisions, and tracking outcomes.[6] Identifying and evaluating validated PRO measures for use at the point of care is thus a priority.

Patient Reported Outcome Measurement Information System (PROMIS) is a National Institutes of Health funded initiative which created item banks of questions in various domains to evaluate PROs across health conditions.[7] PROMIS instruments include computerized adaptive tests (CATs), which utilize item response theory to reduce responder burden and can be easily scored at the point of care. PROMIS has numerous CATs in domains of relevance to SLE patients that have potential utility in measuring PROs in both research and clinical settings. While the feasibility of administering PROMIS CATs to outpatients with rheumatic conditions such as scleroderma and rheumatoid arthritis has been demonstrated, to date there are no studies

evaluating PROMIS CATs in SLE.[8,9] The validation of PROMIS CATs is described in chapter 2, and here we describe the feasibility of administering PROMIS CATs to SLE outpatients participating in the validation study.

METHODS

Population- English speaking adults ages 18 years or older receiving care at the Hospital for Special Surgery (HSS) Lupus Center of Excellence and meeting 4 or more American College of Rheumatology 1997 SLE Criteria were eligible to participate in the validation study.[10] Patients with active malignancy other than non-melanomatous skin cancer and those receiving current dialysis treatment were excluded.

Patient Recruitment and Enrollment- Lupus patients were identified by treating rheumatologists and medical records were screened to confirm eligibility. Patients were approached at the time of their outpatient visit and were invited to participate in a study validating PROMIS CATs in SLE. They were offered the option of completing the web-based surveys on-site or remotely via an emailed study-specific URL. Consenting subjects were registered in Assessment Center (www.assessmentcenter.net), a free secure online research management tool maintained at the Northwestern University Research Data Center, at the time of their visit.

Initial Assessment- Participants completed PROMIS CATs and legacy PRO measures at the time of enrollment either on-site via computer or iPad, or remotely using a device of their choice. Subjects who opted to complete the assessment off-site were emailed a personalized URL to the questionnaire. Subjects who had not completed the assessment by the following day were contacted by phone and/or email with a reminder about the questionnaire and offered assistance with any technical difficulties. Subjects who did not complete the assessment despite the initial reminder were reminded a second time by phone or email. Those who did not complete the assessment after two reminders were re-approached about the questionnaire at a subsequent clinical visit.

Retest Assessment- Participants were contacted by phone and email within one week of enrollment to complete PROMIS CATs a second time for evaluation of retest reliability. Subjects were again reminded up to two times to complete the assessment.

PRO Measures- Fourteen PROMIS CATs were administered as part of the validation study: Physical Function (version 1.2), Mobility (v.1.2), Pain Behavior (v.1.0), Pain Interference (v1.1), Ability to Participate in Social Roles (v2.0), Satisfaction with Social Roles and Activities (v2.0), Fatigue (v1.0), Sleep Disturbance (v1.0), Sleep-Related Impairment (v1.0), Applied Cognition-Abilities (v1.0), Applied Cognition-General Concerns (v1.0), Anger (v1.1),

Anxiety (v1.0), and Depression (v1.0). CATs were programmed to administer enough items to achieve a standard error (precision estimate) of less than or equal to 0.3, corresponding to a reliability of over 0.9 with a minimum of 4 to a maximum of 12 items per CAT. Patients completed two legacy PRO measures: the SF-36 Standard, US version 1.0, and the LupusQoL-US.[11,12] Order of instrument administration was randomized.

Experience Questions- Following completion of the PRO measures in the initial assessment, subjects were asked free response questions about their experience completing the questionnaire. They were asked to comment on what, if any, difficulties they had completing the survey and what they liked about the survey.

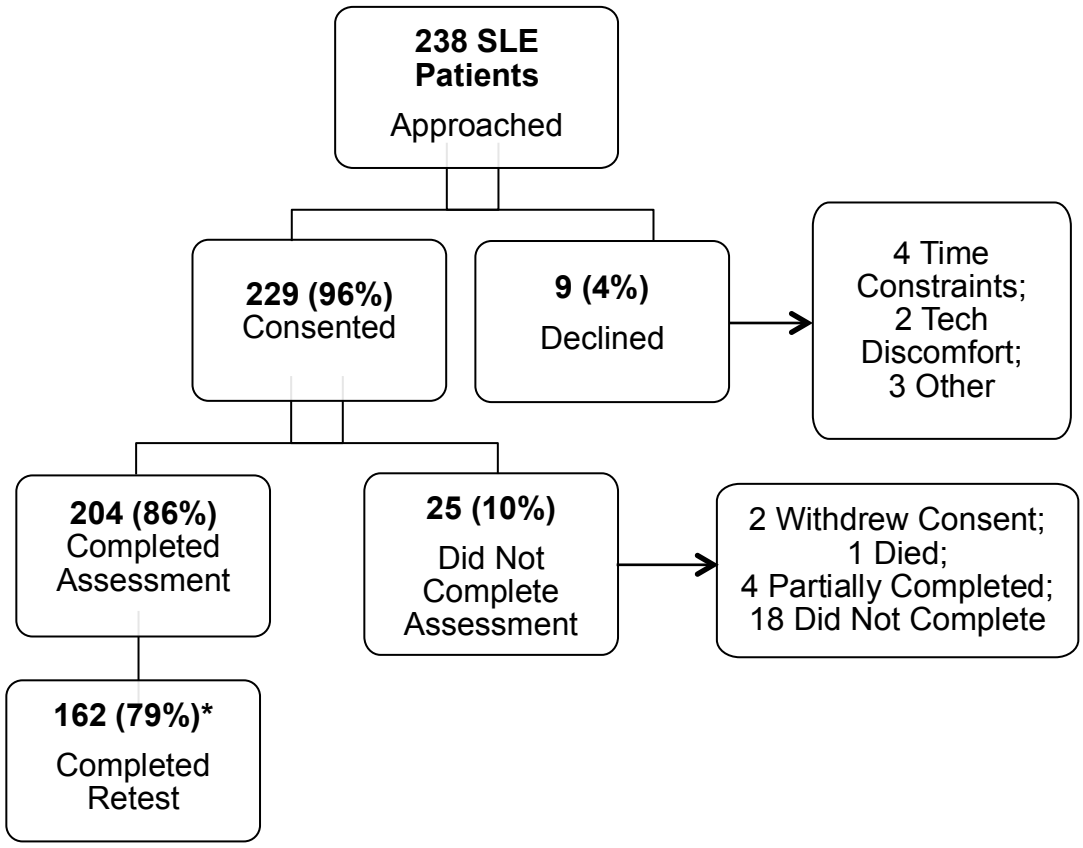
Data Collection- All study questionnaires, including PROMIS CATs and legacy PRO instruments, were administered through Assessment Center. Socio-demographic information including age, sex, race, ethnicity, employment and disability status, education, and insurance type were obtained by patient self-report. Disease duration and medications at the time of enrollment were derived from participants' medical records, while disease activity and damage as measured by SELENA-SLEDAI[13] and SLICC/ACR damage index[14] was provided by the subjects' treating rheumatologists. All information was entered into Assessment Center.

Statistical Analysis- Means and standard deviations were calculated for continuous variables, and frequencies and percentages for categorical variables. Differences in baseline demographics and disease characteristics between participating and non-participating patients were assessed using T-tests and chi-square tests as appropriate. Differences in baseline characteristics were also assessed between patients who completed retest assessments and those who did not. All statistical analyses were performed with SAS version 9.3 (Cary, NC, USA).

The study was reviewed and approved by the HSS Institutional Review Board.

RESULTS

Initial Assessment- Over thirteen months, 238 eligible patients were approached to participate in the study, with 229 patients consenting to participate (Figure 1.1). Among the 9 patients who declined to participate, time constraints was the reason most frequently cited. Of the remaining patients who declined two cited discomfort with computers, one cited a recent injury, one cited discomfort participating in research studies, and one declined because he “did not want to think about lupus symptoms.”



* Of eligible subjects.

Figure 1.1 Study Flow Chart

Two hundred and four subjects went on to complete the initial assessment after consenting, with 164 (80.4%) completing the questionnaire remotely using computers, tablets, or smartphones. Of the 25 patients who consented and did not complete an assessment, 2 withdrew consent from the study, 4 partially completed the questionnaires, and 1 died. Of the 18 remaining subjects who did not complete the assessment many cited time constraints as the reason for not completing the questionnaire during reminder phone

conversations. The majority of consenting subjects who completed the initial assessment did not require reminders (Figure 1.2).

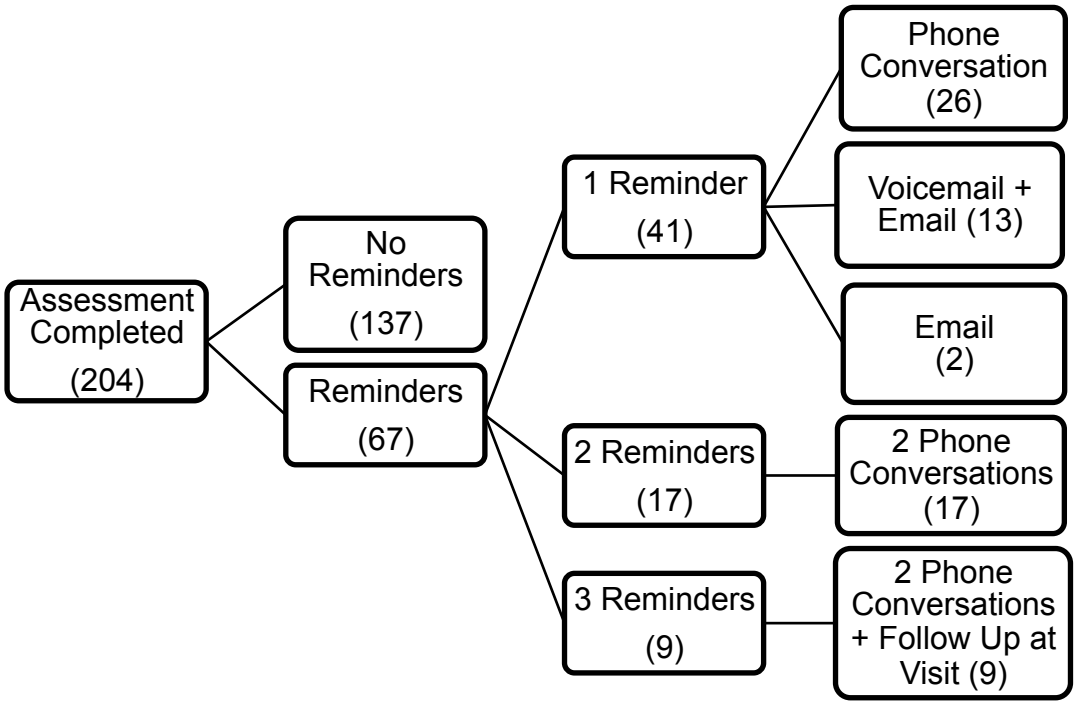


Figure 1.2 Reminders Required for Initial Assessment

Of the 67 subjects (32.8%) who did require reminders, 41 (61.2%) completed the assessment after one reminder, 17 (25.4%) required 2 reminders, and 9 (13.4%) required a third reminder at a subsequent clinical visit. There were no statistically significant differences in demographic or clinical characteristics between subjects who completed the initial assessments and those who did not (Table 1.1), though there was a non-statistically significant trend towards longer disease duration and less frequent renal involvement in non-completers ($p = 0.07$).

Table 1.1 Characteristics of Subjects Completing vs. Not Completing Initial and Retest Assessments

Table 1.1

		Initial			Retest		
Characteristic	**All (n = 238)**	**Completed (n = 204)**	**Did Not Complete (n = 34)**	**p-value**	**Completed (n = 162)**	**Did Not Complete (n = 42)**	**p-value**
Age: mean ± SD years, (range)	40.5± 13.3, (19 - 75)	40.0 ± 13.2, (19 - 73)	43.7 ± 13.9, (24 - 75)	0.13	40.4 ± 12.7, (19 - 73)	48.1 ± 14.7, (21 - 67)	0.30
Female: n (%)	221 (92.9)	189 (92.6)	32 (94.1)	>0.99	151 (93.2)	38 (90.5)	0.52
Race: n (%)				0.64			0.17
White	94 (39.5)	77 (37.7)	17 (50)		65 (40.1)	12 (28.6)	
Black	66 (27.7)	61 (29.9)	5 (14.7)		46 (28.4)	15 (35.7)	
Asian	29 (12.2)	26 (12.8)	3 (8.8)		23 (14.6)	3 (7.1)	
Other	49 (20.5)	40 (19.6)	9 (26.5)		28 (17.3)	12 (28.4)	
Ethnicity: n (%) Hispanic/Latino:	70 (29.4)	58 (28.4)	12 (35.3)	0.48	43 (26.5)	15 (35.7)	**0.04**
Insurance: n (%)				0.88			0.19
Medicaid	86 (36.1)	73 (35.8)	13 (38.2)		53 (32.7)	20 (47.6)	
Medicare	29 (12.2)	21 (10.3)	8 (23.5)		18 (11.1)	3 (7.1)	
Private	123 (51.7)	110 (53.9)	13 (38.2)		91 (56.2)	19 (45.2)	
ACR Classification Criteria: n (%)							
Malar Rash	91 (38.2)	80 (39.2)	11 (32.4)	0.45	60 (37.0)	20 (47.6)	0.21
Discoid Rash	24 (10.1)	21 (10.3)	3 (8.8)	>0.99	16 (9.9)	5 (11.9)	0.78

Table 1.1 Continued

		Initial			Retest		
Characteristic	**All (n = 238)**	**Completed (n = 204)**	**Did Not Complete (n = 34)**	**p-value**	**Completed (n = 162)**	**Did Not Complete (n = 42)**	**p-value**
Ulcers	68 (28.6)	60 (29.4)	8 (23.5)	0.48	50 (30.9)	10 (23.8)	0.37
Arthritis	191 (80.3)	165 (80.9)	26 (76.5)	0.55	133 (82.1)	32 (76.2)	0.39
Serositis	77 (32.3)	67 (32.8)	10 (29.4)	0.59	58 (35.8)	9 (21.4)	0.18
Renal	104 (43.7)	94 (46.1)	10 (29.4)	0.07	68 (42.0)	26 (61.9)	**0.02**
Neurologic	23 (9.6)	19 (9.3)	4 (11.8)	0.70	13 (8.0)	6 (14.3)	0.26
Hematologic	121 (58.8)	101 (49.5)	20 (58.8)	0.22	83 (51.2)	18 (42.9)	0.44
Immunologic	204 (85.7)	178 (87.3)	26 (76.5)	0.15	140 (86.4)	38 (90.5)	0.95
ANA	237 (99.6)	204 (100)	33 (97.1)	0.14	162 (100.0)	42 (100.0)	N/A
Medications: n (%)							
Current Steroid Use	136 (57.4)	118 (58.1)	18 (52.9)	0.40	87 (53.7)	31 (74)	0.07
Current Hydroxychloroquine Use	194 (84.3)	170 (85.9)	24 (75.0)	0.21	138 (85.2)	32 (76)	0.11
Current Immunosuppressive Use	157 (68.3)	138 (69.7)	19 (59.4)	0.65	111 (68.5)	27 (69)	0.12

Table 1.1 Continued

Characteristic	All (n = 238)	Initial			Retest		
		Completed (n = 204)	Did Not Complete (n = 34)	p-value	Completed (n = 162)	Did Not Complete (n = 42)	p-value
Disease Duration: mean ± SD years, (range)	12.8 ± 9.6, (0 - 50)	12.2 ± 8.8, (0 - 48)	16.5 ± 12.9, (1 - 50)	0.07	12.3 ± 8.7, (0 – 48)	11.9 ± 9.3, (0 - 41)	0.79
Physician Global Assessment: mean ± SD, (range) [Range 0 to 3, higher is worse]	NA	0.8 ± 0.6, (0 – 2.8)	NA	NA	0.7 ± 0.6, (0 – 2.6)	0.9 ± 0.7, (0 – 2.8)	0.26
SELENA-SLEDAI: mean ± SD, (range) [Range 0 to 105, higher is worse]	NA	4.2 ± 3.5, (0 – 20)	NA	NA	4.1 ± 3.4, (0 – 16)	4.9 ± 4.0, (0 – 20)	0.17
SELENA-SLEDAI Flare: n (%)	NA	40 (19.6)	NA	NA	27 (16.7)	13 (31)	**0.03**
SLICC: mean ± SD, (range) [Range 0 to 46, higher is worse]	NA	1.2 ± 1.7 (0 – 8)	NA	NA	1.3 ± 1.7 (0 – 6)	1.0 ± 1.4, (0 – 8)	0.52

NA= Not Available (disease activity information not available for non-participants).

Subjects completing the initial assessment had variable levels of education and employment status, with one third of participants reporting receiving disability benefit at the time of enrollment (Table 1.2). Participants self-reported high rates of use of technology, with over 80% reporting regular use of a smartphone and nearly as many (77.7%) reporting daily use of email.

Table 1.2 Additional Characteristics of Participating Subjects (n = 204)

Characteristic	**n (%)**
Education	
High School or less	34 (16.7)
Some College	49 (24.1)
College or Beyond	120 (59.1)
Employment	
Employed- Full Time	74 (36.3)
Employed- Part Time	22 (10.8)
Student	15 (7.4)
On disability	67 (33.0)
Use of Technology	
Computer	148 (72.9)
Tablet	100 (49.3)
Smartphone	172 (84.7)
None of the Above	7 (3.4)
Daily Use of Email	157 (77.7)

Time required to complete the initial assessment varied among subjects- the SF-36 took subjects an average of 6.6 minutes (median = 5.2), the LupusQoL took an average of 5.9 minutes (median = 4.6), and the 14 PROMIS CATs took an average of 11.3 minutes (median = 7.4).

Retest Assessment- One hundred and sixty two subjects (79% of those eligible) completed a retest assessment within one week of the initial assessment. Most of these subjects required a reminder to complete the questionnaire, with 78 (48.1%) requiring one reminder and 51 (31.5%) requiring two reminders (Figure 1.3).

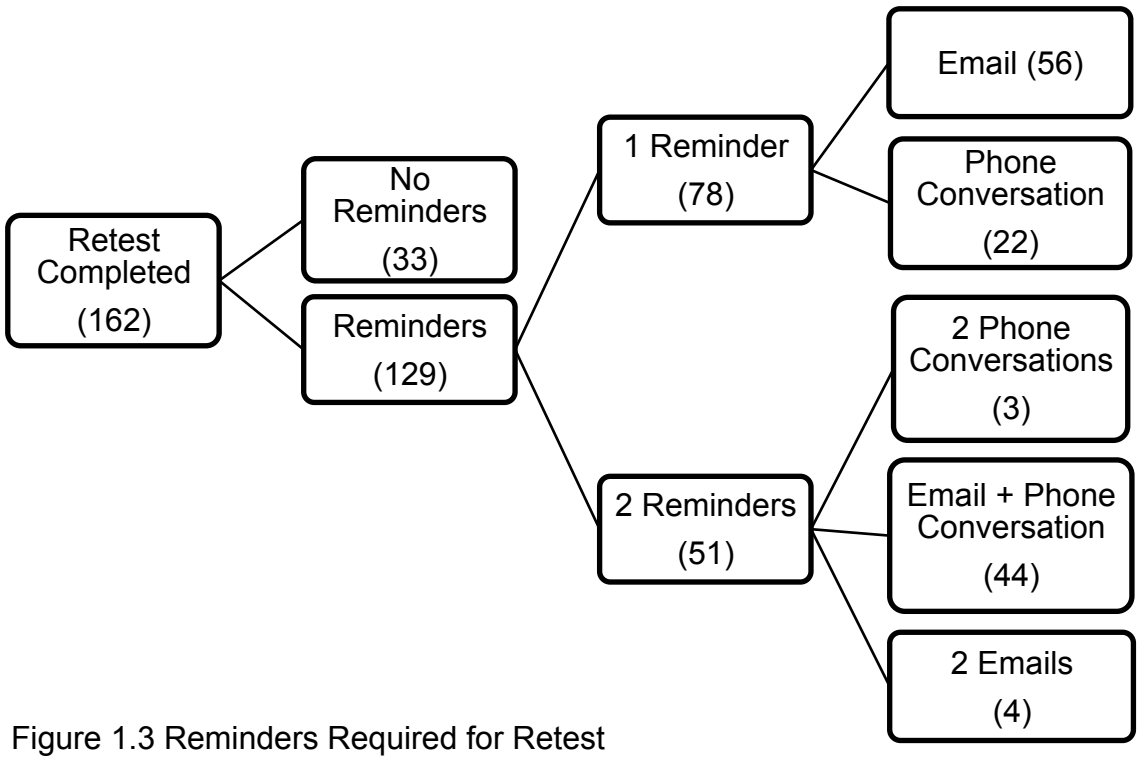


Figure 1.3 Reminders Required for Retest

Patients who completed the retest were more frequently non-Hispanic than those who did not (73.5% vs. 64.3%, p = 0.04) and had less frequent renal involvement by ACR classification criteria than those who did not (42% vs. 61.9%, p = 0.02). They were also less likely to be flaring by SELENA-SLEDAI definition of flare (16.7% vs. 31%, p = 0.03). There were no other statistically significant differences in demographic or clinical characteristics between the two groups (Table 1.1).

Experience- A majority of subjects reported a positive experience completing the questionnaires. Of the 204 subjects completing the initial assessment, 155 commented on aspects of the survey that they liked. Comments related to the content of questions were most frequent, with over 25 subjects praising the concise, clear, detailed, and thorough nature of the questions. Many cited the relevance of the questions- “you get to the heart of the disease symptoms.” Four participants specifically commented on appreciating the inclusion of emotional and mental health questions in the survey. Many also cited the format of the assessment as a positive, with 15 describing it as “easy” or “user-friendly,” 6 appreciating the multiple choice format, and 5 citing the online nature of the survey as a positive, since it could be done on “private time” or at home.

Many participants appreciated the validating and self-reflective aspects of the survey with over 20 subjects specifically commenting on these areas. One subject noted the survey “made me think about how lupus affects me, made me think about how I am really feeling, ”while others stated the survey “showed me things that I did not realize have affected me” and “made me more aware of my lifestyle.” Validation was a recurrent theme with one subject reflecting: “Every question applied so accurately to me personally, there was a strange satisfaction especially after having it take so many years before I was properly diagnosed. That even after so many years that NONE of it is imaginary or overblown alleviates any guilt or doubt that I might now or have ever felt about how I’m feeling or have felt in the past.” Another explained that the survey was “very appropriate for lupus issues that people don’t believe or are hard for the patient to voice to a doctor or a loved one. I often feel stupid

and I am not because I can't remember a word or something is on the tip of my tongue. It is embarrassing and frustrating. It is validating to be asked about it." Others noted "it is refreshing to know that someone cares" and that the survey "gave me a sense of support."

Eighty-two participants commented on challenges completing the survey. The most common difficulty was technical, with 7 subjects reporting error messages from Assessment Center. Six subjects felt the survey was too long with 3 citing the repetitiveness of questions and two noting that a progress bar would be helpful. Four subjects had difficulty understanding the questions, with 4 reporting the variation of time frame in the questions was confusing and 4 noting the questions were at times vague. Five subjects wanted more room to explain their responses or add comments, and one felt that an interview would be more appropriate for gathering this information. Several subjects reported symptoms while completing the survey including pain in the fingers/hands (2), neck/shoulder pain (1), difficulty seeing the screen (1), sleepiness (2), and difficulty concentrating (3). One subject noted difficulty "talking and thinking about lupus- denial is somewhat more comfy."

DISCUSSION

This study is the first to demonstrate the feasibility of administering PROMIS CATs to SLE outpatients. Over 85% of a diverse group of SLE patients approached to participate in the study completed an assessment, the majority without any reminders, and nearly 80% of those went on to complete the retest assessment one week later.

There were no significant demographic differences between patients completing an assessment and those who did not, suggesting that PROMIS CATs are feasible in patients with diverse backgrounds, including those without a college degree (31%), those on disability (33%), and those insured by Medicaid (36%). There were also no significant differences in disease characteristics between those completing the surveys and those who did not, suggesting that patients with mucocutaneous lupus were as likely to complete the survey as those with visceral organ involvement. Arthritis and concentration difficulties were raised by a small number of participants, but overall did not preclude completion of the survey. There were larger percentages of Hispanics and subjects with history of renal involvement among those who did not complete the retest assessment. This will need to be further explored in focused interviews with subjects and/or in larger studies, particularly because this finding was undercut by an opposing trend in subjects completing the initial assessment. Importantly, flaring subjects were somewhat less likely to complete the retest, which raises a concern of feasibility of longitudinal evaluation patients with more active disease. This will need to be more thoroughly investigated in larger studies over a longer period of time.

Subjects had overwhelmingly positive experiences completing the assessment as evidenced by their qualitative feedback. The questions were felt to be relevant, thorough, clear, and concise and the format of online multiple-choice questions was welcomed. Perhaps most importantly, patients commented on the value of completing the assessments, citing the benefits of self-reflection and feelings of validation in being asked about issues relevant to their experience. The promotion of self-reflection and validation through completion

of PROs may be an important mechanism in enhancing patient engagement and satisfaction with their health care experience.

Challenges raised by subjects most commonly related to technical difficulties, as Assessment Center occasionally prematurely ended the survey. Assessment Center will need to be optimized or new vehicles for CAT administration will need to be developed to minimize these difficulties in larger scale use of CATs in research or routine clinical care. The length of the survey, which took subjects over 20 minutes on average, was the second most frequent concern raised. This is unlikely to be an issue in the routine use of PROMIS CATs in SLE, as this study included redundant PROMIS CATs (i.e. physical function and mobility) and both the SF-36 and LupusQoL as part of the validation protocol. Each PROMIS CAT averaged less than one minute in length, which will likely be a tolerable responder burden, though further implementation studies are necessary.

This study has many strengths, including robust rates of enrollment and study completion. The high rate of participation in this study is particularly notable as recruitment of SLE patients in clinical trials is often difficult, with termination of trials due to insufficient enrollment.[15] Reasons for poor enrollment in SLE prevention trials have been studied and include concerns about current health status, trial design, including randomization and medications, and personal factors.[16] The lack of invasive interventions and randomization in this study may have mitigated some of these concerns, but the high participation rate may also reflect SLE patients underlying interest in being asked to share their experience, reinforcing the importance of assessing PROs. Similar findings

were shown in a study evaluating the feasibility of collecting serial electronic PROs (SF-36 and the Hospital Anxiety and Depression scale) in French patients with chronic inflammatory diseases, including 58 patients with SLE.[17] Ninety-six percent of patients approached consented and 89% went on to complete the first assessment with a 13% attrition rate over 6 months. The main reason for non-completion was “being too busy.” Similarly, a longitudinal validation study of the Lupus Impact Tracker, a lupus specific PRO measure, showed a 92% completion rate of retest surveys.[18]

A further strength of this study is the rigorous comparisons of characteristics between responders and non-responders, which has not been performed in previous studies of PROMIS CATs or SLE-specific PROs. This analysis, which showed no significant difference in demographic or clinical characteristics between the two groups, shows that PROMIS CATs can be successfully administered to diverse patients with SLE. The inclusion of qualitative data evaluating the experience of completing PROMIS CATs is a further strength of this study and captures the importance of measuring PROs in validating the patient experience.

There are certain limitations to this study. The feasibility of PROMIS CATs was evaluated in the context of a validation of study in which subjects were administered additional legacy PRO instruments. Thus, completion rates and experience comments do not reflect the experience of completing PROMIS CATs in isolation. Secondly, this study was conducted in a population of English-speakers with relatively high comfort with technology. Further studies

will need to be performed in non-English speakers and in populations with varied familiarity with electronic devices.

Despite these limitations, this study demonstrates the feasibility and importance of administering PROMIS CATs to SLE outpatients. With their SLE-relevant domains and previously proven superior performance characteristics, PROMIS CATs have great potential for capturing PROs in settings related to both research and clinical care. Future studies are needed to evaluate the feasibility of longitudinal collection of PROMIS CATs and their impact on clinical decision-making, patient-provider communication, patient engagement and satisfaction

REFERENCES

1. Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Endpoints: consensus recommendations from OMERACT IV. *Lupus*. 2000;9(5):322-327.

2. FDA, HHS. Guidance for Industry Use in Medical Product Development to Support Labeling Claims Guidance for Industry. *Clin Fed Regist*. 2009;(December):1-39.

3. European Medical League. Guideline on clinical investigation of medicinal products for the treatment of systemic lupus erythematosus and lupus nephritis. 2015:1-16. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500184889.pdf Accessed March 8, 2016.

4. Jolly M, Garris C, Mikolaitis RA, et al. Development and Validation of the Lupus Impact Tracker: a Patient-Completed Tool for Clinical Practice to Assess and Monitor the Impact of Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)*. 2014;66(10):1542-1550.

5. Miedany YM El, Bahlas S, Gaafary M El, et al. Electronic Patient Reported Outcome Measures in Systemic Lupus Erythematosus : A Potential Disease Modifying Impact on the Management of Disease Activity Flares and Damage Accrual. *Arthritis Rheumatol*. 2015;67(suppl 1. http://acrabstracts.org/abstract/electronic-patient-reported- outcome-measures-in-systemic-lupus-erythematosus-a-potential-disease-modifying-impact- on-the-management-of-disease-activity-flares-and-damage-accrual/. Accessed April 10, 2016.

6. Lavallee DC, Chenok KE, Love RM, et al. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. *Health Aff*. 2016;35(4):575-582.

7. Cella D, Riley W, Stone A, et al. Initial Adult Health Item Banks and First Wave Testing of the Patient-Reported Outcomes Measurement Information System (PROMIS) Network: 2005-2008. *J Clin Epidemiol*. 2011;63(11):1179-1194.

8. Khanna D, Maranian P, Rothrock N, et al. Feasibility and Construct Validity of PROMIS and Legacy Instruments in an Academic Scleroderma Clinic-- Analysis from the UCLA Scleroderma Quality of Life Study. *Value Heal*. 2012;15(1):128-134.

9. Bartlett SJ, Orbai A-M, Duncan T, et al. Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PLoS One*. 2015;10(9):e0138543.

10. Hochberg MC. Updating the American College of Rheumatology Revised criteria for the classification of Systemic Lupus Erythematosus. *Arthritis Rheum*. 1997;40:1997.

11. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483.

12. Jolly M, Pickard AS, Wilke C, et al. Lupus-specific health outcome measure for US patients: the LupusQoL-US version. *Ann Rheum Dis*. 2010;69(1):29-33.

13. Petri M, Kim MY, Kalunian KC, et al. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med*. 2005;353(24):2550-2558.

14. Gladman DD. The Development and Initial Validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for Systemic Lupus Erythematosus. *Arthritis Rheum*. 1996;39(3):363-369.

15. Choi CB, Bae SC, Gupta S, Rogers MP, Liang MH. Improving participation in clinical trials of novel therapies: Going back to basics. *Rheum Dis Clin North Am*. 2014;40(3):553-559.

16. Costenbader KH, Brome D, Blanch D, Gall V, Karlson E, Liang MH. Factors determining participation in prevention trials among systemic lupus erythematosus patients: A qualitative study. *Arthritis Rheum Care Res*. 2007;57(1):49-55.

17. Jamilloux Y, Sarabi M, Kerever S, et al. Adherence to online monitoring of patient-reported outcomes by patients with chronic inflammatory diseases: a feasibility study. *Lupus*. 2015:1429-1436.

18. Jolly M, Kosinski M, Garris C, Oglesby AK. Prospective Validation of the Lupus Impact Tracker: a Patient-Completed Tool for Clinical Practice to Evaluate the Impact of Systemic Lupus Erythematosus [Accepted Article]. *Arthritis Rheumatol*. 2016:1-30.

CHAPTER TWO

The Validity of Patient Reported Outcome Measurement Information System (PROMIS) in Systemic Lupus Erythematosus

INTRODUCTION

The accurate measurement of health related quality of life (HRQOL), an important patient reported outcome (PRO), is critical to providing patient-centered care. This is especially important in diseases such as systemic lupus erythematosus (SLE), in which dramatically lower mortality rates have refocused care on minimizing morbidity.[1] It is well known that SLE significantly decreases HRQOL.[2] However, how HRQOL should best be defined and measured is unclear, as physicians and patients have differing perceptions of the impact of SLE: patients focus on functional status whereas physicians focus on laboratory values.[3]

The United States Food and Drug Administration, the European Medical League, and the Outcome Measures in Rheumatology Clinical Trials (OMERACT) group have identified HRQOL as a crucial outcome measure for clinical trials and observational studies in SLE.[4–6] They recommend the use of both generic and disease-specific measures to allow comparisons with healthy individuals while evaluating health utilities that are meaningful to patients.

Numerous generic and disease specific instruments have been validated for the measurement of PROs in SLE, but all have important limitations.[7,8] The Medical Outcomes Survey Short Form 36 (SF-36)[9] is a widely used generic measure in SLE, but has variable longitudinal responsiveness[10–12] and lacks multiple domains of relevance to lupus patients, such as fatigue, sleep, and

cognition.[13–15] The LupusQoL, the most extensively validated SLE-specific instrument, includes several of these SLE-specific domains, including fatigue, body image and planning, but has significant floor and ceiling effects.[16] In addition, both measures are difficult to administer and score at the point of care.

The Patient Reported Outcome Measurement Information System (PROMIS) is a novel psychometrically validated system developed by the National Institutes of Health to efficiently measure PROs in populations with a wide range of chronic diseases.[17] PROMIS instruments increase measurement precision and reduce responder burden relative to traditional instruments as they were developed using item response theory and include computerized adaptive tests (CATs). CATs select the most informative questions from an item bank based on subjects' previous responses, permitting the use of fewer questions per domain with more precision. PROMIS item banks are generic, scored with T scores normalized to the general population in the United States, and include numerous domains of relevance to SLE patients that are not found in the SF-36, including CATs related to fatigue, sleep and cognition.

The performance characteristics of PROMIS CATs have not yet been demonstrated in SLE. This study describes the validity and reliability of 14 PROMIS CATs compared to both the SF-36 and LupusQoL in adult SLE outpatients.

METHODS

Population- English speaking adults ages 18 years or older receiving care at the Hospital for Special Surgery (HSS) Lupus Center of Excellence and meeting 4 or more American College of Rheumatology 1997 SLE Criteria were eligible to participate.[18] Patients with active malignancy, other than non-melanomatous skin cancer, and those currently on dialysis were excluded.

Enrollment- Lupus patients were identified by treating rheumatologists, and medical records were screened to confirm eligibility. Patients were invited to participate in the study and consented at the time of an outpatient visit. Patients were able to complete the web-based surveys on-site during their outpatient visit via computer or iPad with the assistance of a study member. Alternatively, patients were given the option of completing the study questions remotely on computer, tablet, or smartphone via an emailed study-specific URL. Consenting subjects were registered in Assessment Center (www.assessmentcenter.net), a free secure online research management tool maintained at the Northwestern University Research Data Center.

Data Collection- All study questionnaires, including PROMIS CATs and SF-36 and LupusQoL (the latter two to be referred to as “legacy PRO instruments”), were administered through Assessment Center. Socio-demographic information including age, sex, race, ethnicity, employment and disability status, insurance type, and relevant comorbidities were obtained by patient

self-report. Disease duration and medications at the time of enrollment were derived from participants' medical records. Disease activity and damage were assessed by the subject's treating rheumatologist using the Safety of Estrogens in Lupus Erythematosus-National Assessment-Systemic Lupus Erythematosus Disease Activity Index (SELENA-SLEDAI) and the Systemic Lupus International Collaborating Clinics/American College of Rheumatology-Damage Index (SLICC/ACR-DI) respectively.[19,20] SELENA-SLEDAI includes a physician global assessment, ranging from 0 to 3, and the SELENA-SLEDAI score, ranging from 0 to 105. SLICC/ACR-DI scores range from 0 to 46. Higher scores reflect greater disease activity and end organ damage. Half the participants were randomly assigned to complete the PROMIS CATs first, and the other half completed legacy PRO instruments first.

To assess PROMIS CATs test-retest validity, all participants were contacted by telephone or email within one week of enrollment to complete PROMIS CATs a second time. A 7 point Likert scale anchor question was used to identify any changes in patients' disease activity: "Compared to when you last completed this survey, how would you rate the impact of lupus on your general health now? a) Much better now, b) somewhat better now, c) a little better now, d) about the same, e) a little worse now, f) somewhat worse now or g) much worse now."

PRO Measures- PROMIS CATs were selected based on prior focus group studies in which SLE patients identified quality of life domains of critical importance.[14,15,21] Fourteen CATs were administered: Physical Function (version 1.2), Mobility (v.1.2), Pain Behavior (v.1.0), Pain Interference (v1.1), Ability to Participate in Social Roles (v2.0), Satisfaction with Social Roles and Activities (v2.0), Fatigue (v1.0), Sleep Disturbance (v1.0), Sleep-Related Impairment (v1.0), Applied Cognition-Abilities (v1.0), Applied Cognition-General Concerns (v1.0), Anger (v1.1), Anxiety (v1.0), and Depression (v1.0). PROMIS CAT items refer to the seven preceding days, with the exception of items in the physical and social health domains, which do not specify a recall time frame. CATs were programmed to administer enough items to achieve a standard error (precision estimate) of less than or equal to 0.3, corresponding to a reliability > 0.9 with a minimum of 4 to a maximum of 12 items per CAT. Patients completed two legacy PRO measures: the SF-36 Standard, US version 1.0, a frequently utilized generic PRO instrument validated for use in lupus clinical trials, and the LupusQoL-US, an extensively validated lupus specific PRO questionnaire adapted for use in the United States.[9,22] Both legacy instruments referred to a four-week recall period.

PROMIS CATs were scored through Assessment Center using a T score metric, where the mean T score in the US general population is 50 with a standard deviation of 10. Higher T scores reflect more of the trait being measured, so that higher scores for physical and social function are desirable,

whereas higher symptom scores indicate a greater burden of symptoms. The SF-36 is divided into 8 scales, each with a score ranging from 0 to 100, with higher scores reflecting better health related quality of life (HRQOL). Scores are summarized in the Physical Component Summary (PCS) and Mental Component Summary (MCS), which are normalized to the general US population with a mean score of 50 and standard deviation of 10. The LupusQoL contains 34 questions in 8 domains, with scores ranging from 0 to 100, with higher scores indicating better HRQOL.

Statistical Analysis- Means and standard deviations were calculated for continuous variables, and frequencies and percentages for categorical variables. Score distributions are described with means, standard deviations, medians, 25^{th} and 75^{th} percentiles and ranges. Floor and ceiling effects for each instrument were analyzed by calculating the percentage of respondents achieving the minimum and maximum possible scores respectively. Convergent validity between PROMIS CATs and legacy PRO instruments was assessed with Spearman's correlation coefficients (*r*), with coefficients of at least 0.7 indicating good convergent validity. Correlation between PROMIS CATs and disease activity and damage measures were also evaluated with Spearman's *r*. Test-retest reliability was evaluated in participants completing two questionnaires within the seven day time frame who indicated "about the same" in their condition on the anchor question. Agreement between scores for each questionnaire was assessed with interclass correlation coefficient

(ICC). ICCs of at least 0.7 indicate acceptable test-retest reliability. All statistical analyses were performed with SAS version 9.3 (Cary, NC, USA).

The study was reviewed and approved by the HSS Institutional Review Board.

RESULTS

Population- Of 238 eligible SLE patients invited to participate in the study, 229 consented and 204 (86%) completed the questionnaires. Among the 204 participants, 162 (79%) completed the retest within one week. A diverse cohort of patients participated in the study (Table 2.1). Subjects were predominantly female (93%) with mean (SD) age of 40.0 (13.2) years. A majority of participants were non-white (62.3%) and over one quarter identified as Hispanic or Latino (28.4%). Over 40% of subjects were insured by Medicare or Medicaid, with 47.1% working full or part-time and one third reporting receiving disability benefits. With regard to SLE disease characteristics, the average (SD) disease duration was 12.2 (8.8) years, and the most common clinical classification criteria were arthritis (81%), renal manifestations (46%), and photosensitivity (45%). The average (SD) SELENA-SLEDAI score was 4.2 (3.5), indicating mild disease activity, though 19.6% were flaring per SELENA-SLEDAI at the time of assessment. The mean (SD) SLICC/ACR-DI was 1.2 (1.7).

Table 2.1 Characteristics of Participants (n = 204)

Table 2.1 Characteristics of Participants (n = 204)

Characteristic	**Value**
Age: mean ± SD years, (range)	40.0 ± 13.2, (19 - 73)
Female: n (%)	189 (92.6)
Race: n (%)	
White	77 (37.7)
Black	61 (29.9)
Asian	26 (12.8)
Other	40 (19.6)
Ethnicity: Hispanic/Latino: n (%)	58 (28.4)
Insurance: n (%)	
Medicaid	73 (35.8)
Medicare	21 (10.3)
Private	110 (53.9)
Employment: Full or Part-Time: n (%)	96 (47.1)
Disability: n (%)	67 (33.0)
Comorbidities: n (%)	
Anxiety	58 (28.4)
Depression	56 (27.5)
Fibromyalgia	29 (14.2)
ACR Classification Criteria: n (%)	
Malar Rash	80 (39.2)
Discoid Rash	21 (10.3)
Photosensitivity	92 (45.1)
Ulcers	60 (29.4)

Table 2.1 Continued

Characteristic	**Value**
Arthritis	165 (80.9)
Serositis	67 (32.8)
Renal	94 (46.1)
Neurologic	19 (9.2)
Hematologic	101 (49.5)
Immunologic	178 (87.3)
ANA	204 (100)
Medications: n (%)	
Current Steroid Use	118 (58.1)
Current Hydroxychloroquine Use	170 (85.9)
Current Immunosuppressive Use	138 (69.7)
Disease Duration: mean ± SD years, (range)	12.2 ± 8.8, (0 - 48)
Physician Global Assessment: mean ± SD, (range) [Range 0 to 3, higher is worse]	0.8 ± 0.6, (0 – 2.8)
SELENA-SLEDAI: mean ± SD, (range) [Range 0 to 105, higher is worse]	4.2 ± 3.5, (0 – 20)
SELENA-SLEDAI Flare: n (%)	40 (19.6)
SLICC: mean ± SD, (range) [Range 0 to 46, higher is worse]	1.2 ± 1.7 (0 – 8)

PROs- PROMIS CAT and legacy instrument score distributions are shown in Table 2.2. The mean scores were worse than the general population across all PROMIS domains. Mean SF-36 PCS and MCS scores were 1.4 and 0.7 SD worse than the general population. Mean LupusQoL scores across domains

were 10 to 20 points higher than published benchmarks for SLE patients in the U.S.[23] Subjects with self-reported anxiety, depression, or fibromyalgia scored worse than those without these comorbid conditions across all PROMIS domains. Subjects with self-reported fibromyalgia scored one standard deviation worse in the pain interference and ability to participate in social roles CATs than those without (p-value < 0.001). Subjects with self-reported anxiety and depression scored one standard deviation worse in the anger, anxiety, and depression CATs than those without (p-value <0.001).

PROMIS CATs were generally normally distributed, except for pain behavior and fatigue, which had slight positive skews. Similarly, SF-36 scale scores were relatively normally distributed except for the physical function, role physical and role emotional scales, which were positively skewed. All domains in the LupusQoL were positively skewed. The SF-36 had large floor and ceiling effects in the role physical and role emotional scales (23–50%), while the LupusQoL had notable ceiling effects across all domains (6-32%). PROMIS CATs had no floor or ceiling effects.

Table 2.2 PROMIS CAT and Legacy PRO Instrument Score Distribution

(n = 204)

Table 2.2

	Mean	Standard Deviation	Median	25th Percentile	75th Percentile	Min	Percent Min Score	Max	Percent Max Score
PROMIS CAT[a]									
Physical Function	42.6	8.4	41.9	37.7	48.1	24.7	0.5	73.3	1.0
Mobility	42.7	8.4	41.7	36.8	47.9	22.9	0.5	60.0	9.8
Pain Interference	57.1	10.2	58.2	52.2	63.6	38.7	15.2	80.1	0.5
Pain Behavior	56.0	9.3	58.5	53.6	62.2	35.3	12.3	73.8	0.5
Fatigue	58.3	9.5	59.1	51.8	64.0	24.3	0.5	82.9	0.5
Anger	53.9	11.2	54.6	47.4	61.4	28.9	3.4	85.2	0.5
Anxiety	55.8	9.4	57.7	51.1	62.0	32.9	2.5	84.8	0.5
Depression	53.0	10.5	54.3	45.6	60.9	34.2	12.8	78.1	1.0
Ability to Participate In Social Roles	46.8	9.2	45.6	41.7	51.2	21.5	0.5	67.5	6.9
Satisfaction with Social Roles	45.9	9.9	45.8	40.0	52.3	22.0	1.5	68.7	3.9
Sleep Disturbance	56.6	10.7	56.5	51.8	63.0	26.3	2.0	83.8	1.5

Table 2.2 Continued

	Mean	Standard Deviation	Median	25th Percentile	75th Percentile	Min	Percent Min Score	Max	Percent Max Score
Sleep-Related Impairment	57.7	10.4	59.2	51.9	64.3	36.2	1.5	83.1	0.5
Applied Cognition-Abilities	45.8	8.9	43.5	39.5	50.2	26.2	0.5	67.7	5.9
Applied Cognition-Concerns	39.0	11.2	54.6	35.1	44.9	14.5	8.8	69.2	1.0
SF-36β									
Physical Function	57.4	28.7	60.0	35.0	85.0	0.0	2.0	100.0	7.8
Role Physical	35.3	41.4	25.0	0.0	75.0	0.0	49.5	100.0	23.0
Bodily Pain	42.4	24.9	42.0	22.0	62.0	0.0	8.8	100.0	2.9
Vitality	43.2	21.0	45.0	25.0	55.0	0.0	1.0	100.0	0.5
Mental Health	61.7	11.7	60.0	56.0	68.0	8.0	0.5	100.0	0.5
Role Emotional	49.8	43.2	50.0	0.0	100.0	0.0	35.8	100.0	35.3

Table 2.2 Continued

	Mean	Standard Deviation	Median	25th Percentile	75th Percentile	Min	Percent Min Score	Max	Percent Max Score
Social Function	56.9	26.5	56.3	37.5	75.0	0.0	4.9	100.0	11.8
Global Health Perceptions	55.4	13.7	57.0	47.0	65.0	20.0	1.5	97.0	0.5
PCS	36.9	5.5	37.5	32.9	40.9	23.6	0.5	51.2	0.5
MCS	43.0	9.2	44.2	35.7	50.0	19.8	0.5	62.9	0.5
LupusQoL$^{\delta}$									
Physical Health	60.8	26.8	59.4	40.6	82.8	0.0	1.0	100.0	7.8
Pain	57.9	29.5	58.3	33.3	83.3	0.0	4.4	100.0	15.2
Fatigue*	53.2	27.2	56.3	31.3	75.0	0.0	3.0	100.0	5.9
Emotional Health	71.5	25.2	79.2	58.3	91.7	0.0	0.5	100.0	12.7

Table 2.2 Continued

	Mean	Standard Deviation	Median	25th Percentile	75th Percentile	Min	Percent Min Score	Max	Percent Max Score
Planning	65.4	29.4	75.0	50.0	91.7	0.0	3.9	100.0	18.6
Burden to Others	53.1	31.7	50.0	33.3	75.0	0.0	10.3	100.0	12.3
Body Image**	67.0	28.6	75.0	50.0	89.6	0.0	3.3	100.0	20.0
Intimate Relationship	67.1	32.9	75.0	50.0	100.0	0.0	8.3	100.0	31.9

[α]PROMIS CATs are scored by T-score from 0 to 100 (higher signifies more of the measured trait), with a score of 50 corresponding to the general population mean.
[β]SF-36 sub-scales are scored from 0 to 100 (higher signifies better health related quality of life). For the physical component summary (PCS) and mental component summary (MCS), a score of 50 is equivalent to the general population mean.
[δ]LupusQoL domains are scored from 0 to 100 (higher signifies better health related quality of life).
*n = 203
**n = 180

The number of items and time per instrument are shown in Table 2.3. On average PROMIS CATs administered 4 items per domain and the median time per CAT was 32 seconds.

Correlations of PROMIS CATs with Legacy Instruments- Correlations between PROMIS CATs and legacy instruments are shown in Table 2.4. PROMIS physical function and mobility CATs correlated strongly with the physical function domains in the SF-36 and LupusQoL (r = 0.81 - 0.86, p-value < 0.0001), and moderately with the SF-36 PCS (r = 0.49 - 0.52, p-value < 0.0001). Correlations between PROMIS pain interference and legacy instrument pain domains were also strong ($|r|$ = 0.79, p-value < 0.0001). PROMIS fatigue correlated better with the corresponding domain in the LupusQoL ($|r|$ = 0.75, p-value < 0.0001) than with the SF-36 vitality scale ($|r|$ = 0.67, p-value < 0.0001). Similarly, in the domain of mental health, PROMIS anger, anxiety, and depression CATS showed strong correlations with the LupusQoL emotional health domain ($|r|$ = 0.69 – 0.75, p-value < 0.0001), but poor to moderate correlations with all of the SF-36 mental health related scales ($|r|$ = 0.29 – 0.68, p-value < 0.0001). PROMIS social function CATs correlated moderately to strongly with the corresponding domains in the SF-36 and LupusQoL ($|r|$ = 0.55 – 0.75).

Table 2.3 Items and Time per Instrument

Table 2.3

Instrument	Number of Items Administered					Time (minutes)				
	n	mean	median	min	max	n	mean	median	min	max
SF-36	204	36	36	36	36	204	6.6	5.2	0.7	54.8
LupusQoL	204	34	34	34	34	204	5.9	4.6	0.4	48.9
PROMIS CAT										
Physical Function	203	4.3	4	4	12	203	1.0	0.7	0.1	20.2
Mobility	203	5.3	4	4	14	203	0.8	0.6	0.1	4.0
Fatigue	202	4.3	4	4	12	203	0.8	0.5	0.1	14.1
Pain Interference	203	5.4	4	4	12	203	0.8	0.6	0.1	12.0
Pain Behavior	203	5.0	4	4	12	203	0.9	0.6	0.1	7.7
Anger	202	7.1	6	5	12	203	0.9	0.5	0.1	19.9
Anxiety	202	5.1	4	4	12	204	0.5	0.3	0.1	6.8
Depression	202	5.7	4	4	12	204	0.4	0.3	0.1	4.9
Ability to Participate in Social Roles	203	4.8	4	4	12	204	0.9	0.5	0.1	16.9
Satisfaction with Social Roles	203	5.1	4	4	12	204	1.0	0.7	0.1	7.4
Sleep-Related Impairment	202	5.1	4	4	12	204	0.8	0.5	0.1	14.9
Sleep Disturbance	202	5.5	4	4	12	204	0.7	0.5	0.1	5.3
Applied Cognition-Abilities	202	4.7	4	4	12	204	0.8	0.5	0.1	16.3
Applied Cognition-General Concerns	202	5.5	4	4	12	204	1.3	0.7	0.1	20.5
All PROMIS CATs		72.8	58	57	170		11.3	7.4	0.6	170.8

Table 2.4 Correlations Between PROMIS CATs and Legacy PRO Instruments

Table 2.4

PROMIS CAT	**Legacy Instrument/Domain**	**Spearman's *r****
Physical Function	SF-36/Physical Function	0.83
	SF-36/Role Physical	0.65
	SF-36/Physical Component Summary	0.49
	LupusQoL/Physical Health	0.82
Mobility	SF-36/Physical Function	0.86
	SF-36/Role Physical	0.55
	SF-36/Physical Component Summary	0.52
	LupusQoL/Physical Health	0.81
Pain Interference	SF-36/Bodily Pain	0.79
	LupusQoL/Pain	-0.79
Pain Behavior	SF-36/Bodily Pain	0.70
	LupusQoL/Pain	-0.71
Fatigue	SF-36/Vitality	-0.67
	LupusQoL/Fatigue	-0.75
Anger	SF-36/Mental Health	-0.29
	SF-36/Role Emotional	-0.50
	SF-36/Mental Component Summary	-0.58
	LupusQoL/Emotional Health	-0.69
Anxiety	SF-36/Mental Health	-0.35

Table 2.4 Continued

PROMIS CAT	Legacy Instrument/Domain	Spearman's r*
	SF-36/Role Emotional	-0.49
	SF-36/Mental Component Summary	-0.60
	LupusQoL/Emotional Health	-0.75
Depression	SF-36/Mental Health	-0.33
	SF-36/Role Emotional	-0.59
	SF-36/Mental Component Summary	-0.68
	LupusQoL/Emotional Health	-0.75
Ability to Participate in Social Roles	SF-36/Social Function	0.72
	LupusQoL/Planning	0.75
Satisfaction with Social Roles	SF-36/Social Function	0.60
	LupusQoL/Planning	0.55

*p-value < 0.0001

There were no analogous legacy instrument domains to which to compare the four PROMIS CATS evaluating cognition and sleep. However, these CATs showed strong correlations with fatigue. Correlations between fatigue and sleep impairment and applied cognition-concerns were both 0.68, while

correlations between sleep interference and disturbance was 0.62, and applied cognition-abilities and concerns was -0.74 (p-value <0.001 for all).

Correlations of PROMIS CATs with Physician-Derived Measures: Correlations between PROMIS CATs and physician-derived measures of lupus disease activity and disease-related damage are shown in Table 2.5. Correlations were generally weak and non-significant, with the highest correlations observed between CATs in the domains of physical function and pain and the physician global assessment and SLICC/ACR-DI ($|r|$ = 0.27 to 0.37, p-value <0.0001).

Table 2.5 Correlations Between PROMIS CATs and Physician Derived Measures

PROMIS CAT	**Physician Global**	**SELENA-SLEDAI**	**SLICC-ACR DI**
Physical Function	-0.29*	-0.15	-0.27*
Mobility	-0.35*	-0.16	-0.31*
Pain Behavior	0.31*	0.16	0.22
Pain Interference	0.37*	0.22	0.20
Fatigue	0.26	0.19	0.05
Anger	0.26*	0.21	0.07
Anxiety	0.24	0.17	0.12
Depression	0.24	0.17	0.10
Ability to Participate in Social Roles	-0.28*	-0.17	-0.13
Satisfaction with Social Roles	-0.22	-0.14	-0.08

Table 2.5 Continued

PROMIS CAT	Physician Global	SELENA-SLEDAI	SLICC-ACR DI
Cognitive Abilities	-0.24	-0.16	-0.12
Cognitive Concerns	0.22	0.18	0.09
Sleep Disturbance	0.25	0.10	0.11
Sleep-Related Impairment	0.16	0.13	0.10

*p-value ≤ 0.0001.

Reliability: Of the 162 participants who completed PROMIS CATs a second time within 7 days of the initial assessment, 90 reported no change in the impact of lupus on their health. Among these 90 subjects, ICCs were greater than 0.7 across all domains (Table 2.6).

Table 2.6. Test-Retest Reliability of PROMIS CATs (n = 90*)

PROMIS CAT	ICC	SEM
Physical Function	0.86	3.17
Mobility	0.88	2.91
Pain Behavior	0.80	4.37
Pain Interference	0.86	4.01
Fatigue	0.84	3.94
Anger	0.72	5.87
Anxiety	0.78	4.46
Depression	0.86	3.92

Table 2.6 Continued

PROMIS CAT	ICC	SEM
Ability to Participate in Social Roles	0.87	3.52
Satisfaction with Social Roles	0.78	5.17
Cognitive Abilities	0.83	4.21
Cognitive Concerns	0.83	5.73
Sleep Disturbance	0.88	3.83
Sleep-Related Impairment	0.80	4.85

*Number of participants reporting no change in impact of lupus on health at second assessment.

DISCUSSION

This study is the first to demonstrate the validity and reliability of PROMIS CATs in SLE outpatients. PROMIS CATs showed strong correlations with the SF-36 and LupusQoL across analogous domains, supporting the construct validity of the measures. PROMIS CATs showed high test-retest reliability in participants self-reporting no change in the impact of lupus on their health.

Although to our knowledge no prior studies have evaluated PROMIS CATs in adults with SLE, PROMIS short forms, i.e. PROMIS items administered as regular questionnaires, without using computerized adaptive testing, have been evaluated in SLE patients. The PROMIS-29, a 29 question short form composed of items from seven PROMIS item banks (physical function, fatigue,

pain interference, anxiety, depression, sleep disturbance, and satisfaction with social roles), was evaluated in 333 patients with self-reported SLE, recruited from patient advocacy organizations.[24] The PROMIS-29 domain scores were found to correlate with self-reported disease severity. Mahieu et al evaluated the internal consistency of seven PROMIS short forms (physical function, fatigue, pain interference, anxiety, depression, sleep disturbance, and sleep-related impairment) in 123 adults with SLE, finding strong internal consistency among the measures (Cronbach's alpha 0.91 to 0.98).[25] They also showed strong correlations between PROMIS fatigue short form and the self-report Fatigue Severity Scale scores (Spearman's $r = 0.84$, p-value < 0.0001). The authors found that physical activity, measured with an accelerometer, was positively associated with PROMIS physical function ($r = 0.33$, p-value = 0.0003) and negatively associated with pain interference ($r = -0.29$, p-value = 0.001). PROMIS pediatric short forms have also demonstrated construct validity and responsiveness in 100 children with lupus.[23]

In this study, the first to evaluate PROMIS CATs in SLE, participants scored one half standard deviation or more worse than the general population across most PROMIS CATs, with the largest differences in the domains of physical function, mobility, pain interference, fatigue, sleep-related impairment, and applied cognition-concerns. This trend in scores suggests face validity of the instruments given the known lower HRQOL in SLE patients.[26] These findings are also consistent with those of Mahieu et al who reported SLE subjects

scored one half standard deviation worse than the general population in physical function, pain interference, fatigue, sleep disturbance and sleep-related impairment short forms, which are scored using the same T scale metric as CATs.[25] Further, the performance of SLE patients on PROMIS CATs is similar to that of patients with other chronic rheumatologic conditions with similar mean scores across domains reported in validation studies of PROMIS CATs in those populations.[27–29]

Subjects with anxiety, depression and fibromyalgia scored worse on PROMIS CATs than subjects without these comorbidities, with largest difference in pain interference and social abilities for those with fibromyalgia, and in emotional health for those with anxiety or depression. This further supports the fact that PROMIS CATs have construct validity. Since patients with SLE have a high prevalence of comorbid fibromyalgia, anxiety, and depression that can contribute negatively to HRQOL, it is reassuring that PROMIS CATs were, on average, different between these groups.

While prior studies have suggested that PROMIS short forms appear to have good reliability and precision in patients with SLE, this is the first study to compare the performance characteristics of PROMIS instruments to the SF-36 and LupusQoL, two legacy PRO instruments commonly used in clinical research. In this study, in contrast to legacy instruments, PROMIS CATs demonstrated a normal distribution across domains and notably had no floor

or ceiling effects. In lacking floor and ceiling effects, PROMIS CATs are better able to discriminate among individuals at the extremes of the spectrum, and importantly, may be more sensitive to capturing longitudinal change within individuals. The significant floor and ceiling effects observed in the SF-36 and LupusQoL are consistent with score distributions reported in other studies and may contribute to the variable responsiveness of the measures in longitudinal studies.[10,12,30] Further, PROMIS CATs demonstrated high precision and reliability in this study despite the relatively few questions administered in each domain (average of 4 items per domain). The ability of PROMIS CATs to decrease responder burden without compromising precision or reliability is a significant advantage over legacy instruments.

Importantly, PROMIS CATs correlated poorly with physician-derived measures of SLE disease activity, supporting the principle that PROs are an independent outcome measure. In SLE, where defining appropriate outcome measures for clinical trials remains challenging,[31] the patient perspective is particularly important. PROMIS CATs are validated precise tools to capture patient-centered outcomes and complement physician-derived assessments.

Strengths of this study include the large and diverse cohort of subjects with classification criteria confirmed SLE. There was a high rate of participation and subjects had varied disease activity and severity. Rates of fibromyalgia and psychiatric comorbidities in this cohort are similar to those reported in other

SLE cohorts, suggesting our results are generalizable.[32,33] PROMIS CATs were chosen in domains identified as important by SLE patients in prior studies, and a large number of CATs were administered.

This study has certain limitations. In an effort to decrease responder burden, not all PROMIS domains were validated against equivalent gold standard instruments. For example, the sleep and cognition related CATs had no corresponding domains in the SF-36 and LupusQoL, so the construct validity of these measures was inferred by their correlations with other domains. Conversely, PROMIS lacks many domains that are present in the LupusQoL, including body image, planning, and intimate relationships, which are known to be valued by SLE patients. This points to a gap and these item banks will need to be developed. In this study, PROMIS CATs were evaluated in outpatients; the validity and responsiveness in inpatients, who may have worse HRQOL and worse disease activity, may differ and will need to be explored. Importantly, this study is cross-sectional and longitudinal data is needed to evaluate the responsiveness of PROMIS CATs across varying states of disease activity over time.

In conclusion, this study demonstrates the validity and reliability PROMIS CATs in outpatients with SLE. PROMIS CATs provide an easy and accurate method of evaluating patient-centered domains in patients with SLE, and

could be an important metric for measuring relevant patient domains in both clinical research and routine clinical care.

REFERENCES

1. Urowitz MB, Gladman DD, Tom BDM, Ibañez D, Farewell VT. Changing patterns in mortality and disease outcomes for patients with systemic lupus erythematosus. *J Rheumatol*. 2008;35:2152-2158.

2. McElhone K, Abbott J, Teh L-S. A review of health related quality of life in systemic lupus erythematosus. *Lupus*. 2006;15(10):633-643.

3. Alarcón GS, McGwin G, Brooks K, et al. Systemic lupus erythematosus in three ethnic groups. XI. Sources of discrepancy in perception of disease activity: a comparison of physician and patient visual analog scale scores. *Arthritis Rheum*. 2002;47(4):408-413.

4. FDA, HHS. Guidance for Industry Use in Medical Product Development to Support Labeling Claims Guidance for Industry. *Clin Fed Regist*. 2009;(December):1-39.

5. European Medical League. Guideline on clinical investigation of medicinal products for the treatment of systemic lupus erythematosus and lupus nephritis. 2015:1-16. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500184889.pdf Accessed March 8, 2016.

6. Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Endpoints: consensus recommendations from OMERACT IV. *Lupus*. 2000;9(5):322-327.

7. Castelino M, Abbott J, McElhone K, Teh L-S. Comparison of the psychometric properties of health-related quality of life measures used in adults with systemic lupus erythematosus: a review of the literature. *Rheumatology (Oxford)*. 2013;52(4):684-696.

8. Holloway L, Humphrey L, Heron L, et al. Patient-reported outcome measures for systemic lupus erythematosus clinical trials: a review of content validity, face validity and psychometric performance. *Health Qual Life Outcomes*. 2014;12(116):1-14.

9. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483.

10. Kuriya B, Gladman DD, Ibañez D, Urowitz MB. Quality of life over time in patients with systemic lupus erythematosus. *Arthritis Care Res (Hoboken)*. 2008;59(2):181-185.

11. Touma Z, Gladman DD, Ibañez D, Urowitz MB. Is there an advantage over SF-36 with a quality of life measure that is specific to systemic lupus erythematosus? *J Rheumatol*. 2011;38(9):1898-1905.

12. Devilliers H, Amoura Z, Besancenot J-F, et al. Responsiveness of the 36-item Short Form Health Survey and the Lupus Quality of Life questionnaire in SLE. *Rheumatology (Oxford)*. 2015;54(5):940-949.

13. Danoff-Burg S, Friedberg F. Unmet Needs of Patients with Systemic Lupus Erythematosus. *Behav Med*. 2009;35(1):5-13.

14. McElhone K, Abbott J, Gray J, Williams a, Teh L-S. Patient perspective of systemic lupus erythematosus in relation to health-related quality of life concepts: a qualitative study. *Lupus*. 2010;19(14):1640-1647.

15. Jolly M, Pickard AS, Block J a, et al. Disease-specific patient reported outcome tools for systemic lupus erythematosus. *Semin Arthritis Rheum*. 2012;42(1):56-65.

16. McElhone K, Abbott J, Shelmerdine J, et al. Development and validation of a disease-specific health-related quality of life measure, the LupusQol, for adults with systemic lupus erythematosus. *Arthritis Care Res (Hoboken)*. 2007;57(6):972-979.

17. PROMIS: Dynamic Tools to Measure Health Outcomes from the Patient Perspective. http://www.nihpromis.org/. Accessed March 6, 2016.

18. Hochberg MC. Updating the American College of Rheumatology Revised criteria for the classification of Systemic Lupus Erythematosus. *Arthritis Rheum*. 1997;40:1997.

19. Petri M, Kim MY, Kalunian KC, et al. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med*. 2005;353(24):2550-2558.

20. Gladman DD. The Development and Initial Validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for Systemic Lupus Erythematosus. *Arthritis Rheum*. 1996;39(3):363-369.

21. Ow YLM, Thumboo J, Cella D, Cheung YB, Yong Fong K, Wee HL. Domains of health-related quality of life important and relevant to multiethnic English-speaking Asian systemic lupus erythematosus patients: a focus group study. *Arthritis Care Res (Hoboken)*. 2011;63(6):899-908.

22. Jolly M, Pickard AS, Wilke C, et al. Lupus-specific health outcome measure for US patients: the LupusQoL-US version. *Ann Rheum Dis*. 2010;69(1):29-33.

23. Jolly M, Pickard SA, Mikolaitis RA, Rodby RA, Sequeira W, Block JA. LupusQoL-US benchmarks for US patients with systemic lupus erythematosus. *J Rheumatol*. 2010;37(9):1828-1833.

24. Lai J, Kaiser K, Beaumont J, et al. Health-Related Quality of Life in Patients with Systemic Lupus Erythematosus [abstract]. In: *Arthritis & Rheumatology*.Vol 66 S10.; 2014:S1146-S1147. http://acrabstracts.org/abstract/health-related-quality-of-life-in- patients-with-systemic-lupus-erythematosus/.

25. Mahieu MA, Ahn GE, Chmiel JS, et al. Fatigue , patient reported outcomes , and objective measurement of physical activity in systemic lupus erythematosus. *Lupus*. 2016;0(January):1-10.

26. Jolly M. How does quality of life of patients with systemic lupus erythematosus compare with that of other common chronic illnesses? *J Rheumatol*. 2005;32(9):1706-1708.

27. Bartlett SJ, Orbai A-M, Duncan T, et al. Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PLoS One*. 2015;10(9):e0138543.

28. Khanna D, Maranian P, Rothrock N, et al. Feasibility and Construct Validity of PROMIS and Legacy Instruments in an Academic Scleroderma Clinic-- Analysis from the UCLA Scleroderma Quality of Life Study. *Value Heal*. 2012;15(1):128-134.

29. Tomasson G, Farrar JT, Cuthbertson D, et al. Administration of Patient Reported Outcome Measurement Information System (PROMIS) Instruments By Computer Adaptive Testing in Patients with Systemic Vasculitis [abstract]. In: *Arthritis and Rheumatology*.; 2015:67 (suppl 10). http://acrabstracts.org/abstract/administration-of-patient-reported-outcome- measurement-information-system-promis-instruments-by-computer-adaptive-testing-in- patients-with-systemic-vasculitis/. Accessed March 5, 2016.

30. McElhone K, Abbott J, Sutton C, et al. Sensitivity to Change (Responsiveness) and Minimal Important Differences of the LupusQoL in patients with Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)*. 2016. doi:10.1002/acr.

31. Thanou A, Merrill JT. Top 10 things to know about lupus activity measures. *Curr Rheumatol Rep*. 2013;15(6).

32. Palagini L, Mosca M, Tani C, Gemignani A, Mauri M, Bombardieri S. Depression and systemic lupus erythematosus: a systematic review. *Lupus*. 2013;22(5):409-416.

33. Buskila D, Press J, Abu-Shakra M. Fibromyalgia in systemic lupus erythematosus: prevalence and clinical implications. *Clin Rev Allergy Immunol*. 2003;25(1):25-28.